

## **Digital Libraries and Network Publishing with Wide Area Information Servers**

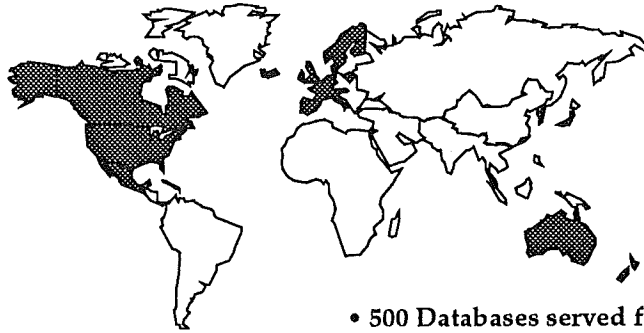
Brewster Kahle  
President  
Wide Area Information Servers, Inc  
Sunsite Japan 12/7/93

## **Trends in Information Access**

- **Wide Area Networks are Driving Changes**
  - Internet is Expanding Very Rapidly
- **Digital Libraries are making Libraries Crucial**
  - Scholarly and Technical Communications is already Digital
- **"Network Publishing" gives Companies an Advantage**

WAIS Inc.

## WAIS on Internet



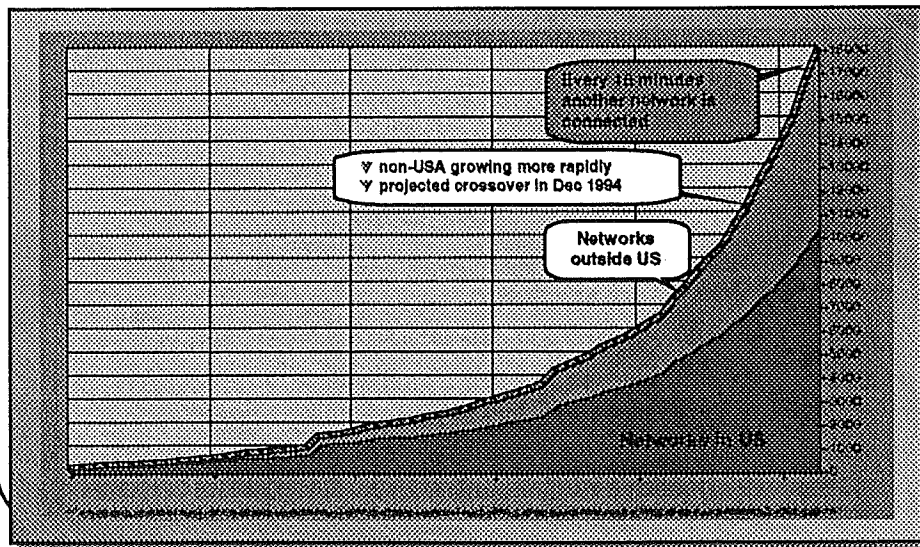
- Over 50,000 users on the Internet
- Users in 32 Countries: Mexico, Singapore, Finland, Australia, ...
- 500 Databases served from 12 Countries: Norway, Canada, UK, ... New database every day.
- VT-100 Interface in California: 240 each day, 4,600 different machines

Sunsite Japan 12/7/93

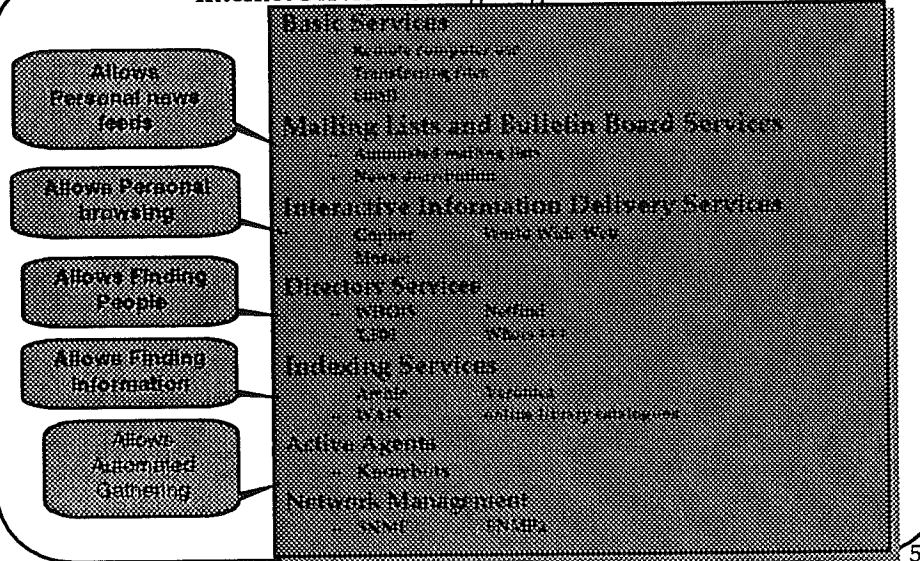
3

WAIS Inc.

## History of Connected IP Networks: The Internet



## Internet Services: Navigating Global Information Space



5

## What I really want...

- My personal information to be accessible
- Published information should find me
- Usable anywhere
- Others can use what I have learned (if I want them to)

What is it?

## **Network Publishing**

**(Finding and retrieving information over wires)**

## **Navigation Techniques: Paper**

- **Alphabetical Listings (dictionary, Encyclopedia)**
- **Indices (back of the book and Readers Guide)**
- **Table of Contents (outlining)**
- **Citation index**
- **"Tree of Knowledge"**
- **Have you read any good books lately?**

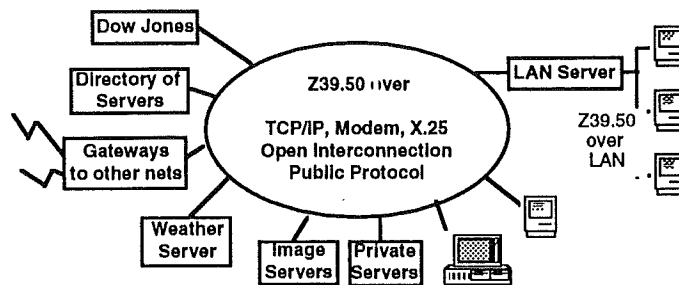
## Navigation Techniques: Computers

- **Hierarchical:**
  - File Folders, FTP, Internet Gopher
- **Filename search:**
  - Unix "find", Mac "find file", Internet Veronica
- **String search:**
  - Grep, ON Location
- **Boolean query systems (AND OR NOT):**
  - Dialog, Mead Data
- **Static Hypertext links (see also pointers):**
  - World Wide Web

## Navigation Techniques: WAIS

- **Natural language questions and relevance feedback as well as Boolean**
  - Question-answer dialog
  - Similar to Newspapers: "More on page 5"
  - Dynamic Hypertext Links
- **2 level search:**
  - Directory of servers
  - Servers themselves

## Wide Area Information Server Architecture



- Point-and-click interface to Worldwide Information
- Client -----> Protocol -----> Server

## WAIS Clients



- Busy 24 hours a day finding information
- Ponder all indications of the preferences of its user
- Gossip with other clients about their discoveries
- Scours the world (within a budget) to find new sources
- Current implementations on PC, Macintosh,
- X Windows, NeXT, dumb terminal (dial-up)

## Networked Back of the Book Index: Searching

- When you know what you are looking for
- Manual indexing ———> full indexing
- Advanced navigation: topic clustering, query analysis, relevance feedback
- Interactive through gigabytes

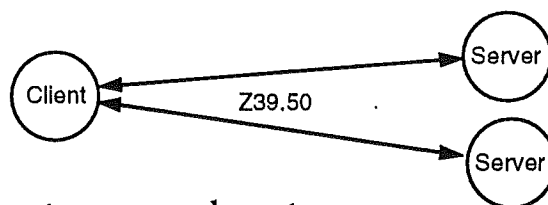
## WAIS Protocol

- Flexible, Extendable, and Standard
- Pieces:
  - Search: (*words, doc\_ids, databases*) returns list of: (*headline, score, doc\_id, types*)
  - Retrieval: (*doc\_id, type, start, end*) returns: *data of specified type*
  - URL: An ISBN for the Electronic Age
  - Server Description Structure for the Directory of Servers
- NISO Z39.50, IETF URL, other international standards

## How Standard Protocol can Provide Security

- Users do not login to server, but search only through application layer protocol (Z39.50).
- Server controls access to data.
- Authentication and Encryption
  - Public Key Encryption
  - Kerberos (WAIS Integration by MCC)
- Enables privacy, billing, and authentication

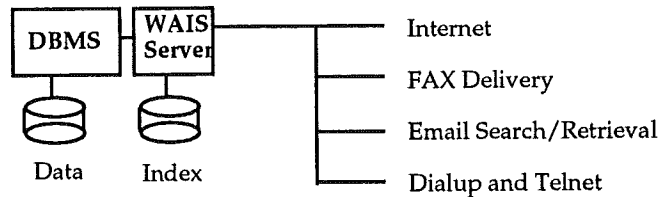
## The WAIS Protocol is WAIS



- Supports any search syntax
- Supports sophisticated clients -- puts intelligence in the user's hands
- Clients can run on any platform
- Multiple servers in a single search
- Retrieve any kind of data: text, graphics, video,...



## WAIS Inc UNIX Server



- 1MB-10GB collections
- Supports 1000s of users
- Easy to install and run
- Boolean, natural Language, Relevance Feedback
- Security and Billing
- Text, images, databases, newsfeeds

## Upcoming Searching Extensions

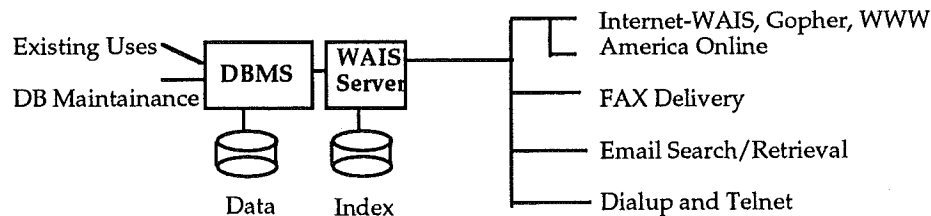
- Integration of TREC/Tipster Research
- Other commercial search engines
- Learning based searches
- Integrated Searching and Browsing (gopher-style)
- International character sets and Unicode
- New protocols (Z39.50-v2, URL, RSA, Search Syntax)

## US Federal Agencies Using WAIS

The OMB Circular A-130 specifically notes: "Emerging standards such as Wide Area Information Servers (using the NISO Z39.50 standard) will be used increasingly to facilitate dissemination of government data..."

- NII: Government Information Locators System
- Department of Energy
- NASA
- US Geological Survey
- Intelligence (Open Source)
- FBI (Terror an Arson)
- Defense Technical Information Center (DTIC)
- US Department of Agriculture
- Environmental Protection Agency
- Securities and Exchange Commission (soon)

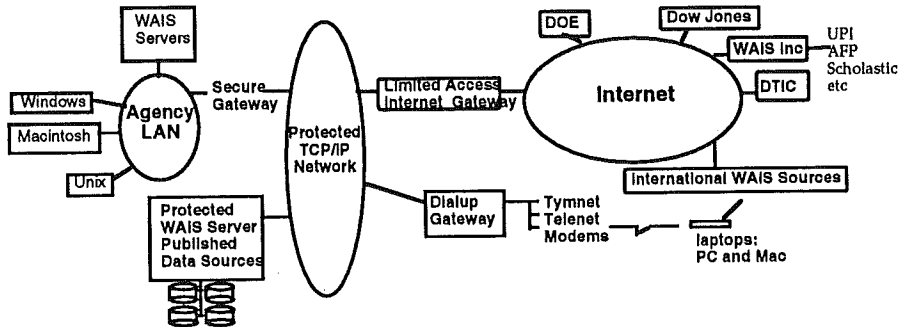
## Defense Technical Information Center WAIS System



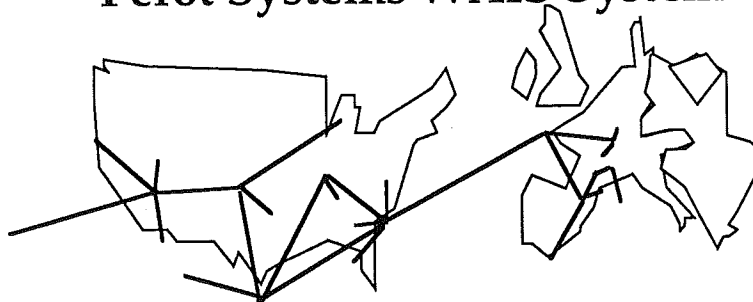
- Project Just Starting (due within 1 year)
- Increased access to Databases and Documents

30 GB

## US Intelligence Agencies Open Source WAIS System



## Perot Systems WAIS System



- Corporate Information
  - Resumes, Company Policies, Presentations, Contracts, Proposals, Phone book
- Published Info
  - CDROM's, Internet resources

## Digital Libraries

- Easy to use card catalog (MARC records)
- Remote use from home or office
- Full text, scanned documents, Pictures

[pegun.law.columbia.e]	columbia-law-library-catalog
[pegun.law.columbia.e]	columbia-spanish-law-catalog
[quake.think.com]	tmc-library

## Network Publishers: Sun Microsystems

- Combined use of Network Tools to Stay in Touch with Customers
- Email Lists to 10,000 users *30,000*
- Archive Downloads of *12,000/day*
- Press Releases, Bug Fixes, Etc

## **A "Network Book" Contains:**

**Table of Contents: Browsing (e.g. Gopher)**

**Pages: With active footnotes, dynamic illustrations (e.g. WWW)**

**Index: Dive into a subject Area (e.g. WAIS)**

**Works over a Network (e.g Internet)**

## **Digital Libraries and Network Publishing**

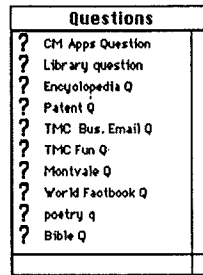
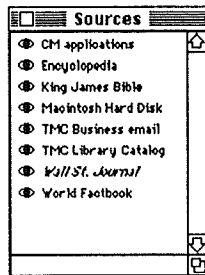
**Large Networks + Too Much Information = A Problem**

**WAIS Can Help Users Find Critical Information**

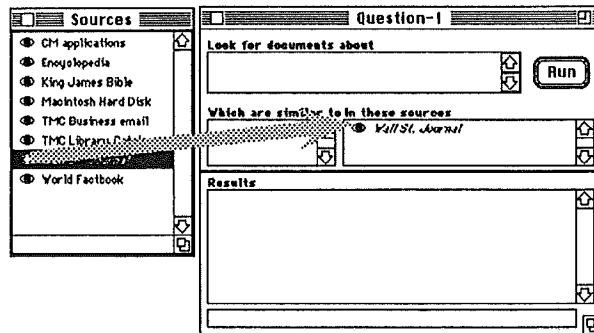
**WAIS Can Help "publishers" Deliver Information to Users**



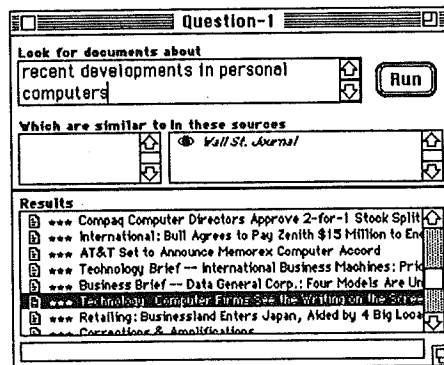
## WAISStation: Active Database Sources, Saved Questions



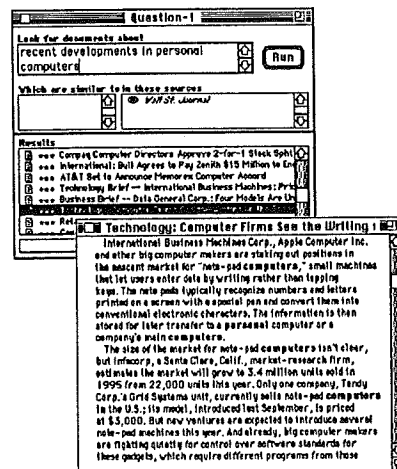
## Select Data Source



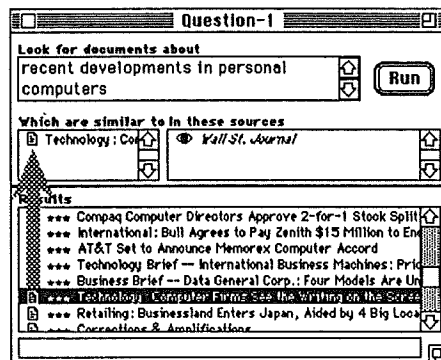
## Run Initial Query



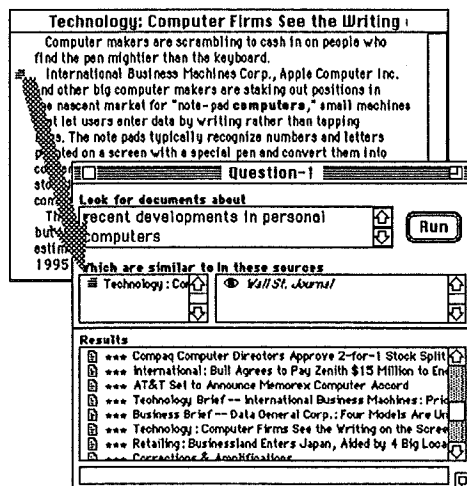
## Click a Headline to Display a Document



## Relevance feedback: "Find me more like this one"



## Relevance Feedback of Paragraph





WAIS Inc.

## "Chaining" of Questions to Follow a Tangent

**Question-1**

Look for documents about  
recent developments in personal computers

Which are similar to in these sources  
Technology: Computers

**Results**

- International: Bull Agrees to Pay Zenith \$15 Million to Exit
- AT&T Set to Announce Memorex Computer Accord
- Technology Brief -- International Business Machines: Price
- Business Brief -- Data General Corp.: Four Models Are Unveiled
- Technology: Computer Firms See the Writing on the Screen

**Question-2**

Look for documents about

Which are similar to in these sources  
Retailing: Business

**Results**

- Retailing: Businessland Enters Japan, Aided by 4 Big Local Firms
- What's News -- Business and Finance
- Technology: Computer Makers Agree on a Standard For Microprocessors
- Inside Track: Businessland Directors Take a Loss And Try to Turn Things Around
- Technology & Health: Businessland To Report Loss For Second Quarter
- Technology: U.S. Computer Maker Takes on NEC on Its Own Terms
- Technology: Computer Firms See the Writing on the Screen